CLAIMS:

1   1.   A system for processing a document having a varying number of pieces of content in

2   a hierarchical document structure, the system comprising:

3   means for identifying an anchor node, the anchor node being a context node of a template for

4   a particular node of content;

5   means for generating a location expression corresponding to the anchor node, the location

6   expression locating one or more pieces of similar content identified by the anchor node; and

7   means for processing the document using the location expression, wherein the location

8   expression is used each time a piece of content corresponding to the anchor node is located in the

9   document so that the document with a varying number of pieces of content underneath the anchor

10   node in the hierarchical document structure are properly processed.

1   2.   The system of Claim 1 further comprising means for identifying an anchor node

2   parent with sibling case where particular nodes of content share the same anchor node and the path

3   expressions for each particular node of content are the same as the anchor node, means for

4   determining the anchors if the anchor node parent with sibling case is identified, means for

5   combining the location expressions of the identified nodes of content into a single location

6   expression for a generalized anchor node, means for determining if the generalized anchor node is a

7   sibling, and means for generating a generalized expression corresponding to the generalized anchor

8   node that locates the content in the particular nodes of content identified.

1   3.   The system of Claim 2 further comprising means for reanchoring the particular nodes

2   of content to a reanchor node if the generalized anchor node is a sibling node and means for

3   determining if the reanchor node is tangled such that the location expression to a piece of content

4   matches more than one piece of content.

1   4.   The system of Claim 2 further comprising means for identifying the lowest node in

2   the hierarchical document structure that has not been generalized and means for generalizing the

3   lowest node before generalizing the nodes that are higher in the hierarchical document structure.

1   5.   The system of Claim 2, wherein the location expression combining means further

2   comprises means for identifying a location expression for each particular node of content, means for

3   determining if there are other nodes of content and means for generating a replacement anchor node

4   if there are no other nodes of content.

6.   The system of Claim 5, wherein the location expression combining means further comprises means for determining if the location expression for the other nodes of content have been generalized, means for generalizing the location expressions of the other nodes of content if they have not been previously generalized and means for identifying the previously generalized location expressions.

7.   The system of Claim 6, wherein the location expression combining means further comprises means for determining if the code associated with the location expression are consistent with each other, means for generalizing each element of a location expression if the code is not consistent and means for generalizing the common elements in the path if the code is consistent.

8.   The system of Claim 3, wherein the means for determining a tangled node further comprises means for determining the anchor nodes in the hierarchical document structure and means for generating replacement nodes for location expressions having the same number of elements if there are no more anchor nodes.

9.   The system of Claim 8, wherein the means for determining a tangled node further comprises means for determining the number of elements in each location expression and means for indexing each location expression according to location, anchor number and element number.

10.   A method for processing a document having a varying number of pieces of content in a hierarchical document structure, the method comprising:

identifying an anchor node, the anchor node being a context node of a template for a particular node of content;

generating a location expression corresponding to the anchor node, the location expression locating one or more pieces of similar content identified by the anchor node; and

processing the document using the location expression, wherein the location expression is used each time a piece of content corresponding to the anchor node is located in the document so that the document with a varying number of pieces of content underneath the anchor node in the hierarchical document structure are properly processed.

11.   The method of Claim 10 further comprising identifying an anchor node parent with sibling case where particular nodes of content share the same anchor node and the path expressions for each particular node of content are the same as the anchor node, determining the anchors if the anchor node parent with sibling case is identified, combining the location expressions of the

5   identified nodes of content into a single location expression for a generalized anchor node,

6   determining if the generalized anchor node is a sibling, and generating a generalized expression

7   corresponding to the generalized anchor node that locates the content in the particular nodes of

8   content identified.

1   12.   The method of Claim 11 further comprising reanchoring the particular nodes of

2   content to a reanchor node if the generalized anchor node is a sibling node and determining if the

3   reanchor node is tangled such that the location expression to a piece of content matches more than

4   one piece of content.

1   13.   The method of Claim 11 further comprising identifying the lowest node in the

2   hierarchical document structure that has not been generalized and generalizing the lowest node

3   before generalizing the nodes that are higher in the hierarchical document structure.

1   14.   The method of Claim 11, wherein the location expression combining further

2   comprises identifying a location expression for each particular node of content, determining if there

3   are other nodes of content and generating a replacement anchor node if there are no other nodes of

4   content.

1   15.   The method of Claim 14, wherein the location expression combining further

2   comprises determining if the location expression for the other nodes of content have been

3   generalized, generalizing the location expressions of the other nodes of content if they have not been

4   previously generalized and identifying the previously generalized location expressions.

1   16.   The method of Claim 15, wherein the location expression combining further

2   comprises determining if the code associated with the location expression are consistent with each

3   other, generalizing each element of a location expression if the code is not consistent and

4   generalizing the common elements in the path if the code is consistent.

1   17.   The method of Claim 12, wherein determining a tangled node further comprises

2   determining the anchor nodes in the hierarchical document structure and generating replacement

3   nodes for location expressions having the same number of elements if there are no more anchor

4   nodes.

1   18.   The method of Claim 17, wherein the determining a tangled node further comprises

2   determining the number of elements in each location expression and indexing each location

3   expression according to location, anchor number and element number.